Effectiveness of Formative Assessment Model on Learning Outcomes

Ilyas[1], Andi Kusumayanti[2]
[1,2]Universitas Islam Negeri Alauddin Makassar
Author Correspondence Email: ilyas.ismail@uin-alauddin.ac.id

ABSTRACT

Assessment of student learning outcomes has not been in line with curriculum expectations as emphasized by government policy. The ineffectiveness of the assessment model is one of the factors that have not optimized student learning outcomes. This study aims to explain the effectiveness of formative assessment on improving student learning outcomes. The research is a form of test on a form of formative testing in the measurement of learning outcomes. This study uses a quantitative approach with experimental research methods. The research sample was 80 students by using two forms of formative assessment analysis, namely Formative Assessment Essay Test Form and Formative Assessment Multiple-choice test form. The results of the analysis of hypothesis testing 1 show that the F-test statistical value above in line B shows that at the $\alpha = 0.05$ level, the value of F count = 10.979, greater than F table = 3.96 so that significant H_0 is rejected, which means that there is a difference in the science learning outcomes of student groups given formative assessment in the form of essay tests with student groups given formative assessment in the form of multiple-choice tests. The results of the analysis of hypothesis 2 testing show that the t-test statistical value of table 5. row [(B=1)] shows that at the $\alpha = 0.05$ level, the t count value = 6.491 is greater than the t table = 1.67 so that Ho is rejected, and it is concluded that the science learning outcomes of student groups given formative assessment in the form of essay tests are higher than those of student groups given formative assessment in the form of multiple choice tests. Thus, formative assessment has significance on student learning outcomes. This study recommends the importance of the form of assessment to be considered in order to improve student learning outcomes.

Keywords: Effectifity, Formative Assessment Type, Learning Achievement

1) INTRODUCTION

Assessment of learning outcomes has been explicitly outlined in the Indonesian Law No. 20 of 2003 concerning the National Education System, but has not run optimally as expected. This is also confirmed through the Regulation of the Minister of Education and Culture (Permendikbud) No. 23 of 2016 concerning Educational Assessment Standards which explains that the assessment of learning

outcomes by educators aims to monitor and evaluate the process, learning progress, and improvement of learning outcomes, but in practice it has not yet obtained significant results. In addition to learning outcomes that are not yet optimal, the form of assessment offered has not gone well. A good assessment format has an impact on good learning outcomes as well (Kunandar, 2015; Salamah, 2018). Similarly, the form of formative assessment has an influence in determining learning outcomes (Ediyanto, 2014; Ismail, 2015; Purnomo, 2013). Based on data from the Head of the Curriculum and Learning Center of the Ministry of Education and Culture, it states that 30 percent of Indonesian teachers experience a mismatch or mismatch between the subjects taught and their educational background. This condition is one of them that makes the quality of education in Indonesia not in accordance with the ideal conditions expected, although many forms of approaches have been applied.

Existing studies on the relationship between assessment and student learning outcomes tend to show three perspectives. First, the implementation of government policies to improve the quality of education (Arwidayanto, Suking, & Sumar, 2018; Bakry, 2010; Solichin, 2015; Sudarto, 2017). Second, the curriculum approach as a medium for improving the quality of student learning (Adang, 2012; Ali, 2013; Alimuddin, 2014; B., 2013), as well as the curriculum development model which has been the subject of much academic debate (Aida, Kusaeri, & Hamdani, 2017; Bungel, 2014; Rusman, 2015; Salamah, 2018) to realize student competencies Third, assessment methods and models that affect student learning outcomes (Farisi, 2011; Hidayani, 2018; Nasbi, 2017; Tarihoran, 2016). From the existing studies, it appears that not much attention has been paid to the use of formative forms in the learning process. In other words, formative forms have not received special attention in answering the problem of suboptimal learning outcomes.

This study is a response to the lack of studies on the effectiveness of formative assessment on student learning outcomes. In line with that, two questions can be formulated: 1) How is the formative assessment process practiced in measuring student learning outcomes? 2). How are the differences in learning outcomes of student groups with formative assessment through essay tests and multiple choice tests. 3). How is the effectiveness of formative assessment on improving student learning scores. These three questions form the basis for the discussion of this study which looks at the differences in student learning outcomes through formative assessment.

This study is built on three arguments; first, the policies and regulations issued by the government related to improving learning outcomes are very firm and clear, but the implementation and practice

in the field have not been well realized. Second, the existing forms of assessment have not been effective, so a formative assessment process is needed as a response to the non-optimal student learning outcomes as expected. Third, providing assessment through formative assessment in the form of essay tests and multiple choice tests has a positive impact on improving student learning outcomes. Thus the form of formative assessment can be used as a solution to the not optimal student learning outcomes in Indonesia.

LITERATURE REVIEW

2.1 Effectiveness

Learning effectiveness is the most important aspect in knowing the quality of education. In measuring the effectiveness of learning, there are various factors that affect the results of learning effectiveness. This achievement is generally influenced by internal factors and external factors. Internal factors consist of (a) Teacher competence, especially language skills, is recognized as an important dimension of teaching competence, because teacher ability is multidimensional (Stehle & Spinath, 2014). Pawlak (2011: 23) suggests that teachers with limited language mastery, in addition to having low self-esteem, low self-satisfaction, learner, and management-satisfaction, and feelings of boredom. Thus, teacher competence is the most important factor in achieving learning effectiveness (Stehle & Spinath, 2014; Mujiburrahman, 2019). (b) Teachers' perception of learning effectiveness. Pre-learning teachers' perception of learning effectiveness is influenced by how teachers teach and students' engagement in learning (Napoles & MacLeod, 2015). Pre-learning teachers should be given opportunities to practice and reflect on teaching experiences related to teaching effectiveness (Park et al., 2014; Rivera-McCutcheon & Scharff Panero, 2014; Napoles & MacLeod, 2015); The way of teaching is an important factor in achieving teacher-to-student learning effectiveness (Napoles & MacLeod, 2015). Teaching methods include the teacher's ability to organize the classroom in order to create an atmosphere conducive to learning (Napoles & MacLeod, 2015).

External factors of learning effectiveness consist of, (a) Population increase significantly affects teaching effectiveness (Potisek et al., 2019). The number of students continues to increase in each class, automatically the teacher will need extra energy to organize the learning process for students; (b) The teacher's preparation process before the lesson (Sadeghi, et al., 2019). The teacher's mastery in a science is very important in its influence to achieve the effectiveness of student learning; © The

process of teacher training; and (d) the learning method chosen by the teacher in the learning process. Learning with good teacher explanation and high student engagement is the most effective learning method (Napoles & MacLeod, 2015). These two factors are essential for young teachers to learn in dealing with complex problems in the classroom (Hannan et al., 2015). Effective school support for novice teachers has the potential to increase teachers' capacity to teach (Napoles & MacLeod, 2015). Teacher support and capacity building considers persistent norms that can be a barrier to creating school improvement (Hannan et al., 2015). Supporting new teacher development can use standardized feedback processes and science improvement methods (Myung & Martinez, 2013; Hannan et al., 2015;). In addition to support from the school, the teacher's ability to self-assessment and reflective power is important in efforts to improve the effectiveness of learning that will be evidenced in student achievement and professional career growth of a teacher (Nielsen, 2014). So that students can get high learning outcomes because student learning outcomes are the main evidence in seeing the effectiveness of learning (McCarthy et al., 2011).

2.2 Formatif Assesment

Formative assessment is defined as a planned process including a set of key pedagogical tools used to obtain ongoing evidence of student understanding that is used by teachers and students to adjust instruction and procedures in lessons (Johnson et al., 2019). According to Bell & Cowie (Decristan et al., 2015) formative assessment is defined as the repeated use of assessment-based information to recognize and respond to student needs to improve learning. Formative assessment will be successful when teachers are able to improve learning practices accompanied by an increase in teacher competence both in material mastery and in classroom management (Dudek et al., 2018). Teachers' competencies play an important role in shaping their assessment practices and have an influence on their ability to teach complex things to be simple (Box et al., 2015). Formative assessment shows that effective teaching and satisfactory student learning outcomes will be achieved if teachers combine specialized teaching practices with high-level processes accompanied by good classroom management (Decristan et al., 2015). Formative assessment has a positive effect on student learning (Decristan et al., 2015). A meta-analysis has shown that formative assessment supports student learning (Kingston & Nash, 2011).

Various factors that influence formative assessment are cognitive activation, a supportive classroom atmosphere, and classroom management by teachers (Decristan et al., 2015). There are

various factors that become obstacles in implementing formative assessment, namely expectations, habits, student dispositions, the pressure felt by teachers to teach subjects oriented to the final exam rather than the development of students' personalities (Box et al., 2015). Research conducted by Rapi (2016) shows that the application of formative assessment appears on student learning outcomes in the field of science. By using formative assessment, it can be seen that the inquiry learning method is more attractive to students and gives positive results not only on grades but also character in students. Educators must be warm and friendly, social and humanist, have a clear vision and commitment to lifelong learning, and execute carefully whatever is planned (Blegur et al., 2017). His expertise in presenting the subject matter can capture the attention of learners. He/she is able to motivate learners (Bhargava & Pathy, 2011; Suyanto & Djihad, 2012) so that there is no forcedness, tightness, fatigue, and laziness during learning activities. Teachers' use of formative assessment is an integral part of classroom practice with the potential to significantly influence student learning outcomes when done well (Kingston & Nash, 2011).

2.3 Learning Achievement

Learning achievement is an important component in analyzing the quality of education (Nath, 2012). According to Rothstein (Nath, 2012) learning is a product not only of classroom activities but also of family, community and peers. Socio-economic, cultural and environmental factors influence student learning and thus school performance. So that in the achievement of learning achievement (learning achievement) is influenced by many things. It is mentioned by Lin & Chen (2015) that student learning achievement is influenced by family conditions. Adolescents who have good learning achievement at school are influenced by good family learning environment conditions (Lin & Chen, 2015). Family conditions and personal abilities are the factors that most influence student learning achievement (Nath, 2012). These include the socioeconomic background of students and their families, and factors related to schools and teachers (Nath, 2012). In addition to family factors, adolescents who do well in the lower stages of learning are likely to have good learning outcomes in the later stages of learning (Lin & Chen, 2015).

The education system is another factor that affects learning achievement. Visual teaching and learning methods have a positive relationship on problem-solving ability and learning achievement (Sung, 2017). Visual tools such as graphic organizers, diagrams, tables and pictures, and mind maps

can help learners to improve their learning achievement (Sung, 2017). This is directly related to students' cognitive abilities (Sung, 2017). In addition to learning methods, principal factors are a recognized component (Leonidas et al., 2010), and research shows that there is a conceptual relationship between principal factors, other school factors, and student learning achievement. These tools may be abstract, such as leadership and vision and may influence student learning achievement through teachers (Sanfo, 2020). In Prasertcharoensuk's research (2015) revealed that life skills have a positive effect on student achievement. Students who are able to communicate to form cooperation with peers and build good relationships with teachers tend to have good learning achievements (Prasertcharoensuk, 2015). The ability to manage stress is also included in the life-skills that influence the process of student achievement (Prasertcharoensuk, 2015).

2). METHOD

The research is a test of a form of formative testing in the measurement of learning outcomes. Two forms of formative assessment were analyzed, namely Formative Assessment of Essay Test Forms and formative assessment of multiple-choice test forms. Ontologically, an essay test is a form of written test, whose composition consists of question items, each of which contains problems and requires student answers through word descriptions that reflect students' thinking abilities. The process of implementing an essay test takes place by: (1) measuring behavior more directly on the learning objectives that have been set, (2) testing students' ability to communicate their ideas in writing, and (3) requiring students to provide answers not just choosing existing answers (Oosterhof & Todorov, 2009; Todorov & Oosterhof, 2011).

A multiple-choice test is a form of question in which answers can be selected from several possible answers that have been provided in a structured manner and tend to be bound (Popham, 2011). Its construction consists of question items and answer choices. Answer choices consist of an answer key and an exception. The answer key must be the correct or most correct answer, while the exception is an incorrect answer, but the trap must work, meaning that students may choose it if they do not master the material.

This study measured the effect of formative assessment on science learning outcomes in elementary school students, after controlling for students' prior knowledge. This research was conducted at SDN 03 am and SDN 05 am in Rawamangun, Kac. Pulo Gadung, East Jakarta. The

implementation time of this study was the odd semester in class V of the 2017/2018 academic year. This research was conducted with a quantitative approach, using experimental research methods. Specifically, this experimental research uses quasi experimental techniques or field experiments. In other words, this study put: (1) the intensity of formative assessment (treatment variable) as the independent treatment variable, (2) science learning outcomes with numerical data scale as the dependent variable (criterion variable), and (3) science prior knowledge with numerical scale as the covariate independent variable. This study used an analysis of covariance (ANKOVA) design with factorial.

The target population in this study were all students of SD 03 and SD 05 morning Rawamangun Village, Pulo Gadung Sub-district, East Jakarta in the 2017/2018 academic year. The target population was all fifth grade students of SD 03 and SD 05 morning. The research sample was 80 students. Kerlinger mentioned that the research sample was taken from the affordable population (Kerlinger & Lee, 2009). Sampling of research both in the experimental class and in the control class was done by simple random technique. There are two data analysis techniques used, namely: (a) descriptive analysis, and (b) inferential analysis, but first the prerequisite analysis test is carried out, namely normality test, homogeneity test, linearity test, covariate regression significance test on the dependent variable, and regression alignment test. The hypothesis tested in this study is about the Effect of Formative Assessment (B) on Natural Science Learning Outcomes (Y) by Controlling students' Initial Knowledge of Natural Science (X).

The homogeneous regression model tested is: $Y_{ijk} = \mu + B_i + X + \varepsilon_{ij}$

Where:

Y_{ijk} = states the observation value of the kth respondent in cell (i, j)

μ = denotes the overall constant parameter

B_i = denotes the effect parameter of the i-th level or treatment of the form factor Form of formative assessment (B)\

X = the score of a single independent variable or covariate

ε_{ij} = denotes a random error

3). RESULTS AND DISCUSSION

1. Descriptive Analysis Results (Formative Assessment Process Practiced in Measuring Student Learning Outcomes)

The formative assessment process applied in measuring students' learning outcomes as shown in the results of descriptive analysis of data regarding the scores of experimental group science learning outcomes, experimental group science initial knowledge, control group science learning outcomes, and control group science initial knowledge. Complete data summarizing the score of science learning outcomes and initial knowledge of science for both groups can be seen in the following table.

Table 1. Recapitulation of Initial Knowledge Score and Student Learning Outcomes

Form of Formative Assessment		X_i	Y_i	X_i	Y_i	X_i	Y_i
B_1	***N***	20	20	20	20	40	40
	$\overline{X}$ / $\overline{Y}$	73,70	82,50	59,30	68,50	65,30	78,75
	S	18,27	8,39	11,55	5,34	14,03	8,72
	Min	30	63	43	72	30	60
	Max	90	90	85	90	85	92
B_2	***N***	20	20	20	20	40	40
	$\overline{X}$ / $\overline{Y}$	61,05	76,10	68,35	81,50	64,30	75,08
	S	10,35	5,25	16,16	6,19	13,50	6,87
	Min	46	76	27	54	30	67
	Max	88	93	82	76	87	97

(Source: data processed by researchers, 2020)

B_1 = *Formative assessment of essay test form*

B_2 = *Multiple choice test formative assessment*

X = *Initial science knowledge of students*

Y = *Science Learning Outcomes of students*

Learning Outcomes of Students Given Formative Assessment in the Form of Essay Test (B_1)

The score of students' science learning outcomes who were given a formative assessment in the form of an essay test obtained the following results: the number of respondents is 40 people, the minimum score is 60, and the maximum score is 92, the empirical score range obtained is 92 - 60 = 32, and the theoretical range is 0 - 100. Furthermore, the data is presented in the form of a frequency distribution table with the number of classes 7, the width of the interval class 5, the average score is 75.08 mode 75.69, median 75.32, and standard deviation 8.72.

Science Learning Outcomes of Students Given Formative Assessment in the Form of Multiple Choice Test (B_2)

The score of students' science learning outcomes who were given a formative assessment in the form of a multiple choice test obtained the following results: the number of respondents is 40 people, the minimum score is 67, and the maximum score is 97, the empirical score range obtained is 97 - 67 = 30, and the theoretical range is 0 - 100. Furthermore, the data is presented in the form of a frequency distribution table with the number of classes 7, the width of the interval class 5, the average score is 78.75 mode 79.10, median 76.50, and standard deviation 6.87.

2. Prerequisite Test (Differences in Learning Outcomes of Student Groups with Formative Assessment through Essay Tests and Multiple Choice Tests)

2.1. Normality test

The normality test is carried out to determine whether the data from each group comes from a normally distributed population or not. In this study, the data normality test was analyzed and tested with the Lilliefors test technique, for statistical hypotheses:

H_0: Data comes from a normally distributed population

H_1: Data does not come from a normally distributed population

The test criteria are: accept H0 if L_o < L table, and reject H_0 if L_o > Ltabel. Normality testing used a significance level of $\alpha = 0.05$, with n = 20, the value of $L_t = 0.190$, and n = 40, the value of $L_t = 0.140$. All groups of science learning outcomes (Y_{ij}) that were tested for normality with the Lilliefors test gave a Lo value or Lilliefors value for the observation results smaller than the Ltabel value, at a significance level of $\alpha = 0.05$ with n = 20, the L_t value = 0.190, and n = 40, the L_t value = 0.140. So

it is concluded that all groups of data on science learning outcomes in this study come from normally distributed populations. Thus, the requirements for data normality can be met.

2.2. Homogeneity Test

Data Homogeneity Test between Groups B1 and B2

The F-test is used to test the homogeneity of data consisting of two groups, namely testing the homogeneity of data between groups B1 and B2. Kadir (2010: 118), The process of analyzing and testing homogeneity can be calculated using the formula

$F=\frac{\text{largest variance(b)}}{\text{smallest variance(k)}}=\frac{S_b^2}{S_k^2}$ to test the hypothesis:

$H_0: \sigma_1^2 = \sigma_2^2$ (the variance of the two groups is homogeneous)

$H_1: \sigma_1^2 \neq \sigma_2^2$ (the variance of the two groups is not homogeneous)

The test is carried out at a significance level of $\alpha = 0.05$ by comparing the Fcount value with the F table value The test criteria are: accept H0 if F count < F table, and reject H0 if F count > F table. Results The results of the analysis with the help of the Microsoft Excel 2007 program obtained the following results.

From the calculation results as in the attachment obtained the value of F count = 1.677 rounded to 1.68. with F table = 1.71. using the significance level $\alpha = 0.05$ and $dk_1 = 39$ and $dk_2 = 39$. Thus F count < F table, so H_0 accepted and concluded between groups B_1 and B_2 have homogeneous variances that are homogeneous.

2.3. Linearity Test

This regression linearity test is conducted to test whether the regression equation model of covariate X on the dependent variable Y is linear or not. This is because inferential statistical testing with ANOVA requires that the regression equation model of covariate X on dependent variable Y must be linear. Testing for linearity is done with the Deviation from Linearity test with the following statistical hypothesis.

H_0: $\hat{Y} = a + bX$ (linear regression model)

H_1: $\hat{Y} \neq a + bX$ (nonlinear regression model)

This linearity test uses a significance level of $\alpha = 0.05$, with the test criteria, namely: accept H0 if F count < F table at the level of $\alpha = 0.05$, and reject H0 if F count > F table at the level of $\alpha = 0.05$.

Based on the results of the analysis, the sig. value in the Deviation from Linearity row is F count = 1.49 < F table = 1.69 at the α = 0.05 level, as well as at the α = 0.01 level, also obtained F count = 1.49 < F table = 2.10 so it can be concluded that H0 is accepted, namely the regression model of the effect of initial knowledge of Natural Science (IPA) on learning outcomes of Natural Science (IPA) is linearly patterned.

2.4. Significance Test of Regression Effect

Testing the significance of the regression effect is intended to determine whether the initial knowledge of science as a covariate variable X has a significant influence or not on science learning outcomes as the dependent variable Y. This test is carried out by testing the significance of the regression coefficient $\hat{Y} = a + bX$ using the F-test. Testing the significance of this regression with the following statistical hypothesis:

H_0: $\beta = 0$

H_1: $\beta \neq 0$.

Testing the significance of the effect of initial science knowledge as covariate variable X on science learning outcomes as dependent variable Y uses a significance level of α = 0.05. The test criteria, namely: accept H0 if F count > F table at the level α = 0.05, and reject H0 if F count < F table at the level α = 0.05.

Table 2. ANOVA for Regression Significance Test

Model	JK	db	RJK	F_{count}	F_{table} *a= 0,05*	F_{table} *a = 0,01*
Regression	1947,698	1	1947,698	44,080	3,96	6,96
Residue	3446,502	78	44,186			
Total	5394,200	79				

(Source: data processed by researchers, 2020)

Based on table 2. obtained sig value in the Regression line is obtained F count = 44.080 > F table = 6.96 at the level of α = 0.01, so that H0 is rejected and it is concluded that the covariate variable X initial knowledge of science has a significant effect on science learning outcomes (Y), then at the

level of α = 0.05, the effect of covariate variable X initial knowledge of science is obtained which is more significant on science learning outcomes (Y), with a value of F count = 44.080 > F table= 3.39.

2.5. Regression Line Alignment Test

Testing the alignment of the regression line is intended to determine the difference in the linear effect of initial science knowledge (X) on science learning outcomes (Y), between the two groups of cells formed by the form of formative assessment factor (B). The statistical hypotheses proposed are as follows.

H_0: $[FS*X]_s = 0$, for all s (regression of all cells aligned)

H_1: Not H_0 (there is regression misalignment).

The hypothesis testing process above is carried out with the F-test for the source of variance FS*X using the significance level α = 0.05 with the test criteria: accept H0 if the value of Fcount < Ftable, and reject H0 if the value of Fcount > Ftable.

Table 3. Analysis Results for Regression Line Alignment Test Based on Data (FS,X,Y)

Source of Variance	JK res	Db	RJK	F_{count}	F_{table}	
					α = 0,05	α = 0,05
Corrected Model	3406,498(a)	7	486,643	17,628		
Intercept	8652,249	1	8652,249	313,408		
X	812,399	1	812,399	29,427		
FS	68,537	3	22,846	0,828		
FS * X	**74,454**	**3**	**24,818**	**0,899**	**2,72**	**4,04**
Error	1987,702	72	27,607			
Total	484036,000	80				
Corrected Total	5394,200	79				

(Source: data processed by researchers, 2020)

Based on table 3. in the FS*X row above, the value of Fhitung = 0, 899 < Ftabel = 2.72 at the α = 0.05 level so that H0 is accepted, when compared to α = 0.01, then Fhitung = 0, 899 < Ftabel = 4.04 so that Ho is more significantly accepted and it is concluded that there is no significant difference in the slope of the regression line (slopes) of all cell factors or research sample groups

3. Research Hypothesis Testing (Effectiveness of Formative Assessment on Improving Student Learning Outcomes)

3.1. Main Hypothesis Testing

The analysis model used is analysis of covariance (ANKOVA), using the Univariate GLM procedure with the aim of testing the effect of the main factor (main effect) on science learning outcomes by controlling students' initial science knowledge.

Table 4. F-Test Statistics of A B A*B on Science Learning Outcome Y by Controlling X

Source of Variant	JK_{res}	Db	RJK	F_{hitung}	F_{tabel}	
					α = 0,05	α = 0,01
Corrected Model	3332,043(a)	4	833,011	30,296		
Intercept	10854,687	1	10854,687	394,782		
X	763,443	1	763,443	27,766		
B	**301,872**	**1**	**301,872**	**10,979**	**3,96**	**6,96**
Fallacy	2062,157	75	27,495			
Total	484036,000	80				
Corrected Total	5394,200	79				

(Source: data processed by researchers, 2020)

Based on the results of the analysis of hypothesis 1 testing, it shows that the F-test statistical value above in row B shows that at the level of α = 0.05, the value of Fcount = 10.979, greater than Ftable = 3.96 so that significant H0 is rejected, which means that there is a difference in the science learning outcomes of student groups given formative assessment in the form of essay tests with student groups given formative assessment in the form of multiple choice tests by controlling students' initial knowledge of science. Furthermore, the one-party hypothesis tested is: It is suspected that the science learning outcomes of students who are given formative assessments in the form of essay tests, with groups of students who are given formative assessments in the form of multiple choice tests after controlling students' initial knowledge of science.

The statistical hypothesis is as follows:

Ho : $\mu B_1 \leq \mu B_1$

$H_1 : \mu B_1 > \mu B_1$

Table 5. T-test Statistics on Y learning outcomes between All Levels of Factor B for Each Level of Factor B by Controlling X

Parameters	B	*Std. Error*	t_{count}	t_{table}	
				α = 0,05	α = 0,01
Intercept	64,739	3,623	17,871		
X	,246	,047	5,269		
[B=1]	-3,290	1,748	-1,883		
[B=2]	0(a)	.	.		
[B=1]	**11,076**	**1,706**	**6,491**	**1,67**	**2,39**
*[A=2] * [B=1]*	0(a)	.	.		
*[A=2] * [B=2]*	0(a)	.	.		

(Source: Data processed by researchers, 2020)

Based on the results of the analysis of hypothesis 2 testing, it shows that the t-test statistical value of table 5. row [(B=1)] shows that at the level of α = 0.05, the value of tcount = 6.491 is greater than ttable = 1.67 so that Ho is rejected, and it is concluded that the science learning outcomes of student groups given formative assessment in the form of essay tests are higher than those of student groups given formative assessment in the form of multiple choice tests, after controlling for students' science knowledge.

DISCUSSION

Based on the results of the description previously described, formative assessment by giving essay tests is able to improve students' critical attitudes in explaining freely. This opens up opportunities for the development of critical thinking ability assessment formats according to the conditions and needs of students (Zubaidah, Corebima, & Mistianah, 2015). On the basis of the hypothesis put forward in this study, it can be explained through theoretical studies as follows; when students are taught with formative assessment in the form of essay tests, students will be happy and feel challenged to complete it. Students can devote all their abilities to explain and describe in detail

the things that are asked. Students can also construct answers verbally and in writing by using their reasoning power to answer questions well. Thus, formative assessment based on the essay test form produces optimal science learning outcomes, compared to formative assessment in the form of multiple choice tests, because multiple choice formative tests do not require students to answer in detail, systematically, and do not require students to construct answers verbally or in writing. Written tests in the form of description questions (essays) are often chosen because they can better measure the extent to which students understand the material being taught (Hayatin & Department, 2015; Tarhadi, Kartono, & Yumiati, 2007). Multiple choice tests require students to apply the formulas they have learned by choosing one answer from several available alternatives. Based on the description above, that for the group of students who used formative assessment in the form of essay tests was higher than the group of students who were given formative assessment in the form of multiple choice tests after controlling for students' prior knowledge of science.

Effectiveness of Assessment of Learning Outcomes through Essay Tests

One of the important indicators in the assessment of learning outcomes is how a learner is able to explore his/her potential to the fullest without any limitations. Giving an essay test is the right method in forcing/conditioning learners to prepare their competence to the maximum (Abbas & Herdi, 2018; Afoan, Sepe, & Djalo, 2016). The use of Essay Tests in learning evaluation plays a very important role in being able to determine the success of learning. The Essay Test form used in evaluation really helps students to be able to maximize all the knowledge they have in writing to answer the questions asked. For the effectiveness of essay tests in practice must consider three things, among others: (a) Questions that show all their mastery of the required knowledge, (b) Questions that require better answers from students than others, (c) questions that are made, can be as complete and specific as possible, without disturbing the purpose of measuring achievement results.

The use of essay tests is one form of written test, whose composition consists of question items, each of which contains problems and requires student answers through word descriptions that reflect students' thinking abilities (Sinambela, 2016; Sutrisno, 2015). The process of implementing essay tests takes place by measuring competencies more directly on the learning objectives that have been set. Giving essay tests is intended to fully develop student responses. Essay tests are also used to fully develop students' ability to provide answers or responses to questions given. Essay tests not only

require the ability to remember and apply a concept but also require sharpness of analysis and interpretation in answering essay tests (Ellery, 2008). Essay tests also require students to recall, interpret, or analyze both questions and answers, rather than simply identifying, preparing alternative answers as happens in multiple choice tests.

Essay encourages students to express ideas in writing

The process of conducting an essay test is to encourage students' ability (competence) in communicating their ideas in writing, and also requires students to provide answers not only to have existing answers, but also he must ideally describe the competencies he has correctly and systematically. Essay tests are also a form of test that requires answers in the form of relatively long descriptions, students do not have answers but provide answers with the freedom to express ideas in their own words, and also explain that essay tests require students to express answers and state in writing (Therrien, Hughes, Kapelski, & Mokhtari, 2009).

Essay tests are a form of evaluation where answer choices are not provided, and students must answer with sentences, so that it can train students in conveying information verbally, besides that essay exams also demand a better understanding of a science and can be used to measure the level of student understanding of a science in more depth (Juhansar, Pabbajah, & Karim, 2016). Tests with essay systems remain the choice of teachers to evaluate the level of students' understanding abilities despite the fact that it is not easy to provide an objective assessment of student answers (Leckie & Baird, 2011; Sulistyo, Saptono, & Asshidiq, 2015). Teachers need a lot of time to check essay answers, the more the number of exams and the number of students who take the exam, the more the number of exams corrected by the teacher.

Essay Encouraging a Commitment to Learning in Students

Formative assessment in the form of an essay (description), is a method that can be understood that essay tests can further increase students' commitment and motivation to learn, and do not provide many opportunities for students to speculate and take chances. It also encourages students to express their ideas in writing and expressively (Farra, Somasundaran, & Burstein, 2015). Essay tests have the advantage of building commitment to learning which is characterized by two perspectives: (1) the power of the questions to measure complex learning outcomes and high cognitive levels, and (2) giving children the opportunity to compose answers according to their own way of thinking. Thus, essay

tests are considered an effective learning process in order to train students to reflect on themselves through verbal narratives.

This skill is very important in the life of society because individuals in society not only make choices of alternatives but must use other alternatives that are more useful. Although essay tests have a significant influence in building students' commitment to learning, they have some weaknesses. These weaknesses tend to be in five aspects: (1) essay tests have low reliability, (2) require a lot of time to check test results (3) limited material, (4) subjective scoring, less consistent, and less reliable, and also (5) require a long time in scoring. However, essay tests can practice conveying information verbally, essay tests also demand a better understanding and commitment to knowledge and can be used to measure the level of human understanding of knowledge in more depth. In line with that, essay tests also provide opportunities for students to think critically according to their skills (Juhansar et al., 2016).

4). CONCLUSION

The effectiveness of formative assessment on student learning outcomes turns out to show a positive impact, as shown in this study which states that the assessment of students who are given formative assessment in the form of essay tests with groups of students who are given formative assessment in the form of multiple choice tests. It turns out that it is empirically tested by the data, this statement is reinforced by the acquisition of descriptive statistical scores, that specifically in the group of students given the formative assessment of the essay test form of 82.50 is higher than the average science learning outcomes of the group of students given the formative assessment of the multiple choice test form of 76.10.

Based on the results of hypothesis testing built in this study, it can be concluded that the results of hypothesis testing are significantly proven where essay tests used as formative assessments have an impact on student learning outcomes. Similarly, based on other findings in this study as a result of the research, it illustrates that specifically in the group of students who were given formative assessment in the form of essay tests compared to the provision of formative assessment in the form of multiple choice tests, this can be seen from the significant difference in learning outcomes. Thus it can be argued that in science learning, it is very appropriate to be given to groups of students who are given

formative assessments in the form of essay tests. In other words, formative assessment has a significant effect on improving student learning outcomes.

This study still has limitations because the object used as study material is still small scale, as well as the analysis used is descriptive and imperential quantitative, so further studies are needed with a larger scale of objects with more comprehensive qualitative analysis. This study is an introduction to formative assessment as a practical offer in order to improve student learning outcomes. Thus, formative assessment is expected to be the answer to the problems of education that have not been optimal in evaluating student learning outcomes in Indonesia.

REFERENCES

104, P. M. N. (2014). Lampiran Tentang Penilaian Hasil Belajar oleh Pendidik pada Pendidikan Dasar dan Pendidikan Menengah. *Pedoman Evaluasi Kurikulum*.

Abbas, M. F. F., & Herdi, H. (2018). Analisis Kemampuan Menulis Esai Argumentatif Mahasiswa Semester 5 Prodi Pendidikan Bahasa Inggris FKIP UNILAK Tahun Akademik 2017/2018. *Lectura : Jurnal Pendidikan*. https://doi.org/10.31849/lectura.v9i1.918

Adang, S. (2012). Kurikulum dan Pembelajaran. *Bandung: Yayasan Kesuma Karya*.

Afoan, M. Y., Sepe, F., & Djalo, A. (2016). Efektivitas Penerapan Model Pembelajaran Think Pair Share (TPS) Terhadap Hasil Belajar Dan Aktivitas Siswa Pada Materi Sistem Pernapasan Manusia. *Teori, Penelitian, Dan Pengembangan*.

Aida, N., Kusaeri, K., & Hamdani, S. (2017). Karakteristik Instrumen Penilaian Hasil Belajar Matematika Ranah Kognitif yang Dikembangkan Mengacu pada Model PISA. *Suska Journal of Mathematics Education*. https://doi.org/10.24014/sjme.v3i2.3897

Ali, M. (2013). IMPLEMENTASI KURIKULUM PENDIDIKAN NASIONAL 2013. *Jurnal Pedagogi*.

Alimuddin. (2014). Penilaian dalam kurikulum 2013. *Seminar NAsional Pendidikan Karakter*.

Arwidayanto, A., Suking, A., & Sumar, W. T. (2018). Analisis kebijakan pendidikan. In *Journal of Chemical Information and Modeling*. https://doi.org/10.1017/CBO9781107415324.004.

B., M. (2013). Kurikulum Berbasis Kompetensi. *Paradigma*.

Bakry, A. (2010). Kebijakan Pendidikan sebagai Kebijakan Publik. *Jurnal MEDTEK*.

Bhargava, A., and Pathy, M. 2011. Perception of student teacher about teaching competencies.

American International Journal of Contemporary Research.Vol. 1, No. 1: 77-81.

Box, C., Skoog, G., & Dabbs, J. M. (2015). A Case Study of Teacher Personal Practice Assessment Theories and Complexities of Implementing Formative Assessment. American Educational Research Journal, 52(5), 956–983. doi:10.3102/0002831215587754.

Blegur, J., Wasak, M. R. P., & Manu, L. (2017). Penilaian formatif peserta didik atas kompetensi pendidik dalam proses pembelajaran. Satya Widya, 33(2), 117-127.

Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Büttner, G., Fauth, B., … Hardy, I. (2015). Embedded Formative Assessment and Classroom Process Quality. American Educational Research Journal, 52(6), 1133–1159. doi:10.3102/0002831215596412.

Dudek, C. M., Reddy, L. A., Lewka, A., Hua, A. N., & Fabiano, G. A. (2018). Improving Universal Classroom Practices Through Teacher Formative Assessment and Coaching. Assessment for Effective Intervention, 153450841877291. doi:10.1177/1534508418772919

Bungel, M. F. (2014). Penerapan Model Pembelajaran Problem Based Learning Untuk Meningkatkan Hasil Belajar Siswa Kelas Viii Smp Negeri 4 Palu Pada Materi Prisma. *Pendidikan Matematika Tadulako*.

Ediyanto. (2014). Pengembangan Model Penilaian Formatif Berbasis Web untuk Meningkatkan Pemahaman Konsep Fisika Siswa. *Jurnal Pendidikan Sains*. https://doi.org/10.17977/jps.v2i2.4499

Ellery, K. (2008). Assessment for learning: A case study using feedback effectively in an essay-style test. *Assessment and Evaluation in Higher Education*. https://doi.org/10.1080/02602930701562981

Farisi, M. I. (2011). Guru sebagai pengembang kurikulum pendidikan multikultur: Tinjauan dari Perspektif Epistemologis. *Temu Ilmiah Nasional Guru III*.

Farra, N., Somasundaran, S., & Burstein, J. (2015). *Scoring Persuasive Essays Using Opinions and their Targets*. https://doi.org/10.3115/v1/w15-0608

Hayatin, N., & Jurusan, K. (2015). Sistem Penilaian Jawaban Essay Otomatis Berdasarkan Nilai Kedekatan Kalimat. *Seminar Teknologi Dan Rekayasa (SENTRA)*.

Hidayani, M. (2018). MODEL PENGEMBANGAN KURIKULUM. *At-Ta'lim : Media Informasi Pendidikan Islam*. https://doi.org/10.29300/attalim.v16i2.845

Ismail, M. I. (2015). Pengaruh Intensitas Penilaian Formatif terhadap Hasil Belajar IPA dengan Mengontrol Pengetahuan Awal Siswa. *Formatif: Jurnal Ilmiah Pendidikan MIPA*. https://doi.org/10.30998/formatif.v2i1.82

Juhansar, Pabbajah, M., & Karim, S. A. (2016). The Implementation of Higher Order Thinking Skills at Universitas Teknologi Yogyakarta in Indonesia : Opportunities and Challenges. *Proceeding of The International Conference on Education and Higher Order Thinking Skills*.

Kerlinger, F. N., & Lee, H. B. (2009). The Foundation of Behavioural Research. In *Foundation of Behavioral Research*.

Kunandar. (2015). Penilaian Autentik (Penilaian Hasil Belajar Pesrta Didik Berdasarkan Kurikulum 2013. In *Jurnal Evaluasi Pendidikan*. https://doi.org/10.21009/JEP.022.05

Leckie, G., & Baird, J. A. (2011). Rater effects on essay scoring: A multilevel analysis of severity drift, central tendency, and rater experience. *Journal of Educational Measurement*. https://doi.org/10.1111/j.1745-3984.2011.00152.x

Mujiburrahman, M. (2019). State and Religion in Aceh: The Competences of Religious Education Teachers (Referring to ACT 14, 2005). *Al-Albab*. https://doi.org/10.24260/alalbab.v8i1.1330

Nasbi, I. (2017). Manajemen kurikulum: *Idaarah*.

Oosterhof, N. N., & Todorov, A. (2009). Shared Perceptual Basis of Emotional Expressions and Trustworthiness Impressions From Faces. *Emotion*. https://doi.org/10.1037/a0014520

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *Teacher Educator*. https://doi.org/10.1080/08878730.2011.605048

Purnomo, Y. W. (2013). Keefektifan penilaian formatif terhadap hasil belajar matematika mahasiswa ditinjau dari motivsi belajar. *Prosiding Seminar Nasional Matematika Dan Pendidikan Matematika*.

Hannan, M., Russell, J. L., Takahashi, S., & Park, S. (2015). Using Improvement Science to Better Support Beginning Teachers. Journal of Teacher Education, 66(5), 494–508. doi:10.1177/0022487115602126

Johnson, C. C., Sondergeld, T. A., & Walton, J. B. (2019). A Study of the Implementation of Formative Assessment in Three Large Urban Districts. American Educational Research Journal, 000283121984234. doi:10.3102/0002831219842347

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. Educational Measurement: Issues and Practice, 30, 28–37.doi:10.1111/j.1745-3992.2011.00220.x

Leonidas, K., Bert, C., Panayiotis, A., & Demetris, D. (2010). A Synthesis of Studies Searching for School Factors: Implications for Theory and Research. British Educational Research Journal, 36(5), 807–830. https://doi.org/10.1080/01411920903165603.

Lin, C.-Y., & Chen, C.-H. (2015). Early Childhood Family Learning Environment's Influence on Adolescent Learning Achievement in Taiwan. Australasian Journal of Early Childhood, 40(2), 20–29. doi:10.1177/183693911504000204

McCarthy, M. A., Niederjohn, D. M., & Bosack, T. N. (2011). Embedded Assessment:A Measure of Student Learning and Teaching Effectiveness. Teaching of Psychology, 38(2), 78–82. doi:10.1177/0098628311401590

Myung, J., & Martinez, K. (2013). Strategies for enhancing the impact of post-observation feedback for teachers (Report). Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Napoles, J., & MacLeod, R. B. (2015). Influences of Teacher Delivery, Student Engagement, and Observation Focus on Preservice Teachers' Perceptions of Teaching Effectiveness. Journal of Music Teacher Education, 25(3), 53–64. doi:10.1177/1057083715580436

Nath, S. R. (2012). Factors Influencing Primary Students' Learning Achievement in Bangladesh. Research in Education, 88(1), 50–63. doi:10.7227/rie.88.1.5

Nielsen, L. D. (2014). Teacher Evaluation:Archiving Teaching Effectiveness. Music Educators Journal, 101(1), 63–69. doi:10.1177/0027432114536736

Park, S., Takahashi, S., & White, T. (2014). Developing an effective teacher feedback system (Report). Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Pawlak M (2011) The role of in-service training for language teachers in the domain of language competence. Glottodid Actica 38(1): 21–30. http://dx.doi.org/10.14746/gl.2011.38.2

Potisek, N. M., Page, L., Narayan, A., McNeal-Trice, K., & Steiner, M. J. (2019). The Association Between Pediatric Faculty Factors and Resident Physician Ratings of Teaching Effectiveness. Global Pediatric Health, 6, 2333794X1882299. doi:10.1177/2333794x18822996

Rapi, N. K. (2016). Pengaruh model pembelajaran dan jenis penilaian formatif terhadap hasil belajar IPA siswa SMPN. Cakrawala Pendidikan, (1), 85877

Rivera-McCutcheon, R., & Scharff Panero, N. (2014). Lowinference transcripts in peer coaching: A promising tool for school improvement. International Journal of Mentoring and Coaching in Education, 3(1), 86-101.

Sadeghi, K., Richards, J. C., & Ghaderi, F. (2019). Perceived versus Measured Teaching Effectiveness: Does Teacher Proficiency Matter? RELC Journal, 003368821984593. doi:10.1177/0033688219845933

Stehle, S., & Spinath, B. (2014). Intended Course Objectives and Perception of Teaching Effectiveness. Psychology Learning & Teaching, 13(3), 205–217. doi:10.2304/plat.2014.13.3.205

Suyanto dan Djihad, A.(2012). Bagaimana Menjadi Calon Guru dan Guru Professional (cet. 1). Yogyakarta: Multi Pressindo.

Prasertcharoensuk, T., Somprach, K., & Ngang, T. K. (2015). Influence of Teacher Competency Factors and Students' Life Skills on Learning Achievement. Procedia - Social and Behavioral Sciences, 186, 566–572. doi:10.1016/j.sbspro.2015.04.021

Sung, E. (2017). The influence of visualization tendency on problem-solving ability and learning achievement of primary school students in South Korea. Thinking Skills and Creativity, 26, 168–175. doi:10.1016/j.tsc.2017.10.007

Sanfo, J.-B. M. B. (2020). A three-level hierarchical linear model analysis of the effect of school principals' factors on primary school students' learning achievements in Burkina Faso. International Journal of Educational Research, 100, 101531. doi:10.1016/j.ijer.2020.101531

Rusman. (2015). Pembelajaran Tematik Terpadu : Teori, Praktik dan Penilaian. *Rajawali Press*. https://doi.org/10.1016/j.matdes.2019.107759

Salamah, U. (2018). PENJAMINAN MUTU PENILAIAN PENDIDIKAN. *Journal EVALUASI*. https://doi.org/10.32478/evaluasi.v2i1.79

Sinambela, I. T. (2016). Tes Esai Pemetaan Konsep sebagai Alat Ukur dalam Belajar Bermakna. *Jurnal Ilmu Pendidikan*. https://doi.org/10.17977/JIP.V4I1.1066

Solichin, M. (2015). Implementasi Kebijakan Pendidikan Dan Peran Birokrasi. *Jurnal Studi Islam Oktober*. https://doi.org/1978-306X

Sudarto, Z. (2017). Implementasi Kebijakan Penyelenggaraan Pendidikan Inklusif. *Jurnal Pendidikan (Teori Dan Praktik)*. https://doi.org/10.26740/jp.v1n1.p97-106

Sulistyo, M. E., Saptono, R., & Asshidiq, A. (2015). PENILAIAN UJIAN BERTIPE ESSAY MENGGUNAKAN METODE TEXT SIMILARITY. *Telematika*. https://doi.org/10.31315/telematika.v12i2.1422

Sutrisno. (2015). Penilaian pembelajaran tematik pada kurikulum 2013 di sekolah dasar. *Jurnal Sekolah Dasar*.

Tarhadi, Kartono, & Yumiati. (2007). Penggunaan Tes Uraian Dibandingkan Dengan Tes Pilihan Ganda Terstruktur Dan Tes Pilihan Ganda Biasa. *Jurnal Pendidikan*.

Tarihoran, N. (2016). Pengembangan Kurikulum. In *Journal of Chemical Information and Modeling*. https://doi.org/10.1017/CBO9781107415324.004

Therrien, W. J., Hughes, C., Kapelski, C., & Mokhtari, K. (2009). Effectiveness of a test-taking strategy on achievement in essay tests for students with learning disabilities. *Journal of Learning Disabilities*. https://doi.org/10.1177/0022219408326218

Todorov, A., & Oosterhof, N. (2011). Modeling social perception of faces. *IEEE Signal Processing Magazine*. https://doi.org/10.1109/MSP.2010.940006

Zubaidah, S., Corebima, A. D., & Mistianah. (2015). Asesmen Berpikir Kritis Terintegrasi Tes Essay. *Symposium on Biology Education. ISBN: 978-602-72412-0-6*.